Entropy Ordering of Context Bits

One attractive property of locality sensitive hashing (LSH) is that its binary outputs are intentionally designed to have relatively balanced statistics. For random projection LSH, each hyperplane ideally divides the input distribution roughly in half, producing bits with probabilities close to 50/50. Consequently:

- Each LSH bit tends to carry a similar amount of information (approximately one bit of entropy).
- There is usually no preferred order in which the bits are consumed.
- Decision trees, Atlas neural networks, or other context-conditioned models can examine the bits in essentially any sequence with little impact on their overall effectiveness.

This is not generally true for arbitrary binary context sources.

Examples include:

- Binarized metadata.
- Thresholded continuous variables.
- Text-derived binary indicators.
- Rule-based feature detectors.

These often exhibit highly unequal entropy. Some bits may be almost always zero, while others may be close to balanced. Their mutual information with the prediction target can also vary dramatically.

For these context sources it is often advantageous to order the bits before consumption.

Useful ordering criteria include:

- Shannon entropy.
- Mutual information with the target.
- Information gain.
- Margin or discriminative power.
- Correlation with previously consumed bits.

Decision trees already exploit this principle by selecting the highest-value split first. The same idea naturally extends to Atlas neural networks: high-information context bits can be consumed early to establish a coarse partition of the input space, while progressively weaker or more specialized bits refine the selected region.

Viewed this way, LSH offers an unusually convenient property. Rather than requiring feature ranking or ordering, it produces a collection of approximately equal-information geometric samples. This simplifies both implementation and analysis, since the model need not determine an optimal ordering before training.

One subtle extension is that while the *marginal entropy* of LSH bits is approximately equal, their **conditional entropy** decreases as more bits are observed because nearby hyperplanes can become correlated on a particular dataset. Even so, compared with most engineered binary features, LSH bits remain remarkably uniform, which is one reason they make such a clean context signal for Atlas-style architectures.